91602
new 82-10

COMPLIMENTARY COPY
Please Credit Ontario Hydro Research if
Reproduced in Whole or in Part

ontario hydro
research division

To Mr. H. N. Issac
Manager - Nuclear Materials Management

DISTRIBUTION OF THE ACTIVE LIQUID WASTE DISCHARGE CONCENTRATION

A. H. C. Chan

Mathematician
Reliability and Statistics Section
Operation Research Department

In assessing the proposal for removing the on-line liquid effluent monitor (LEM) from the Darlington NGS-A design, it was required to estimate the probability that the concentration of β-γ emitters in the active liquid waste (ALW) tank discharges exceeds a specified level. To achieve this, it was necessary to know the underlying distribution of the ALW discharge concentration. Since the distribution could only be estimated from the historical data, it was also important to provide the confidence interval for the estimated probability. Using the ALW discharge records of Pickering and Bruce NGS-A, it was found that the log-normal distribution provided the best fit for the data. The frequency of the tank concentration exceeding the specified level of 24000μCi/m³ was estimated to be 1 in 200,000 years at Bruce NGS-A and 1 in 100,000 years at Pickering. The 99% upper confidence limits are 1 in 2777 years and 1 in 77 years, respectively.

job	file	date	report no.
740672-110-40162	809.7	March 28, 1985	85-112-K

1. INTRODUCTION

In [1], a proposal for removing the liquid effluent monitor from the Darlington NGS-A design was examined. Part of the argument for the removal was the assumed low risk of accidental discharge of ALW. It is required that short-term variations in the concentration of β-γ emitters in waters available to the public should not exceed 10 MPC. For Darlington NGS-A, under the most restrictive operating conditions, the concentration of β-γ emitters in the discharged ALW that would cause this limit to be exceeded is 24000μCi/m^3. For convenience, this value will be called the *Critical Concentration* and the event that a discharge of an ALW tank having a concentration equal or higher than the *Critical Concentration* is called the *Reference Event*. The main objective of this report is to estimate the probability of occurrence of the *Reference Event*, based on the available operational data from Bruce NGS A and Pickering NGS A.

In order to estimate this probability, one needs to know the underlying distribution of the concentration. Since this distribution can only be estimated from the available data, the resulting probability will have a degree of uncertainty associated with it. Consequently, a confidence interval for the estimated probability must also be calculated.

In summary, the following three tasks are required:

1. Estimate the distribution of the ALW discharge concentration.
2. Using this distribution, estimate the probability of occurrence of the *Reference Event*.
3. Determine the upper confidence limit for the above estimate.

2. DISTRIBUTION OF THE DISCHARGE CONCENTRATION

Operating data from the period of 1981 to 1984 on ALW discharges from Pickering A and Bruce A were supplied by Mr. I. Benovich of the Radioactivity Management & Environmental Protection Department. After examining the preliminary statistics (mean, standard deviation, maximum, minimum, etc.) of the data, it was apparent that all the recorded concentrations of ALW discharges were orders of magnitude lower than the *Critical Concentration*. This means that the estimation must involve an extrapolation of the data far beyond the range of the available observations. Hence, it is necessary to determine the form of the distribution

which dictates the formulation of the extrapolations. This can be done by using the method of statistical hypothesis testing.

Hypothesis testing must first select a null hypothesis to be tested. Clearly, there are many possible distributions, eg the normal, exponential, Weibull, extreme value, to name a few. Surely, one cannot test them all. Hence, an objective method to select the null hypothesis is needed. The principle of maximum entropy (see [2]), widely used in the discipline of information theory, is applicable here.

In this section, a brief discussion of the maximum entropy principle is given. Some results are stated without proofs; these can be found in Reference 2. Using these results, the distribution of discharge concentration can be estimated.

2.1 Maximum Entropy Principle

The term "entropy" is defined as the expected information content of a message. A message is a set of possible events that may occur as well as their probabilities of occurrence. The information contained in an event is proportional to the logarithm of the reciprocal of the probability of that event [2].

Suppose that a set of samples $x_1, x_2, \ldots, x_n$ has been observed and that some properties of the population have been derived from the samples. Then, the maximum entropy principle suggests that a distribution that maximizes the entropy under these constraints (the derived properties) is a good candidate to be considered as the distribution for the population. In other words, this distribution is an objective choice for the null hypothesis in a goodness-of-fit test. If the null hypothesis is accepted, the distribution will be considered as the distribution for the population.

The following are a few examples of distributions generated from the maximum entropy principle:

a. Suppose that x_k, $k=1,\ldots,n$ are samples selected from a population with an unknown density function $f(x)$. If the sample mean and the sample standard deviation provide a good indication of the true mean and standard deviation of the population, then it can be assumed that the first two moments of $f(x)$ exist. Now, the distribution which maximizes the entropy, under the constraints that the first two moments exist, is

$$\phi(x) = (2\pi\sigma^2)^{-1/2}\exp\{-(x-\mu)^2/2\sigma^2\}$$

which is called the normal distribution. If $\mu = 0$ and $\sigma = 1$, then ϕ is called the standard normal distribution.

b. Suppose that the samples are all positive and it can be assumed that so is the population. Furthermore, from the sample, there is a good indication of the population mean. Then, it is the exponential distribution that maximizes the entropy.

2.2 The Distribution of the Concentration

During the initial tabulation of the data, it was apparent that the sample mean was not stable because one extremely large concentration can double its value. Therefore, it would seem that the distribution of the concentration would not fit the normal or the exponential distributions, where a stable sample mean is required.

Fortunately, the log transform of the samples has a stable mean. The variance of the samples is also stable. This means that normal distribution maximizes the entropy of the log transforms of the samples. In other words, the concentrations of the ALW discharges could have a log-normal distribution. Thus the log-normal distribution is selected to be the null hypothesis.

In Appendix I, the procedure to test the log-normal hypothesis is outlined. Figures 1 and 2 show that the cumulative sample distributions for Bruce A and Pickering NGS's are similar to that of a log-normal distribution. In fact, the differences between the sample and the theoretical distributions are within the bounds of acceptance. Hence, the log-normal hypothesis is accepted.

3. THE ESTIMATE OF THE REFERENCE EVENT PROBABILITY AND ITS CONFIDENCE INTERVAL

Let q_n denote the estimate of the probability of occurrence of the *Reference Event*. Since by the above result, the discharge concentration has a log-normal distribution, the estimate is

$$q_n = \phi[(10.086-\mu_n)/\sigma_n]$$

where μ_n and σ_n are the sample mean and standard deviation of the

log transform; and Φ is the cumulative distribution function of the standard normal random variable. The upper confidence limit of the q_n, denoted by U, is defined by

$$P\{ q_n < U \} = \alpha,$$

for some pre-selected α value. This limit is determined by the uncertainty of the parameters μ_n and σ_n. Details of its derivation are given in Appendix II.

The results of the estimates and their upper confidence limits (with $\alpha = 0.99$) are summarized in the following table.

Station	Samp Size	Samp* Mean	μ_n	Samp* S.D.	σ_n	q_n	U (α=0.99)
Bruce	2085	45.691	3.1337	78.618	1.1732	$< 10^{-8}$ **	7.2×10^{-7}
Pickering	2907	13.139	1.0445	59.304	1.7499	$< 10^{-8}$ **	1.4×10^{-8}

* The unit is in $\mu Ci/m^3$
** This value is less than the approximation error

4. CONCLUSION

Considering that the average number of ALW discharges in Bruce A is less than 500 tanks per year, and that the corresponding number for Pickering A is 1000 tanks per year, the following conclusions can be derived.

1. The estimated frequency of occurrence of the *Reference Event* is less than:

   a. once in 200,000 years at Bruce NGS-A;
   b. once in 100,000 years at Pickering NGS.

2. The upper confidence limits in the above table can be interpreted as follows: there is only a one percent chance that the frequency of occurrence of the *Reference Event* will be more than:

a. once in 2777 years at Bruce NGS-A;
b. once in 71 years at Pickering NGS.

The critical concentration of 24000 $\mu Ci/m^3$ was derived as the most restrictive value of the β-γ emitter concentrations in the ALW of Darlington NGS-A. Based on the similarity of station design between Darlington A and Bruce A, it may be concluded that the expected frequency of an ALW discharge exceeding the critical concentration in Darlington is less than once in every 2777 years.

The method of analysis developed in this report can be used to derive the expected frequency of occurrence of a similar event at any of the Ontario Hydro nuclear generating stations.

Approved by:

J. G. Cassan
Manager
Operations Research Dept.
(Ext 122+7615)

Prepared by:

A. H. Chan
Mathematician
Operations Research Dept.
(Ext 122+6173)

REFERENCE

[1] J. Villagran, "Assessment of the Need for an On Line Liquid Effluent Monitor at Darlington NGS", Nuclear Materials Management Department, August 27, 1983, File No. 967-NK-38-67978.

[2] Clough, D. J., "Application of the Principle of Maximizing Entropy in the Formulation of Hypotheses", CORS Journal, vol. 2, no. 2, 1964.

[3] Gibbons, J. D., "Nonparametric Statistical Inference", McGraw-Hill, N. Y., 1971.

APPENDIX I

Testing of Hypothesis

The hypothesis to be tested is that the ALW discharge tank concentration has a log-normal distribution; or the log of the concentration has a normal distribution. This hypothesis is the so-called distributional assumption. There are many procedures devised for testing the normality assumption (eg the Shaprio-Wilk test). Unfortunately, they are mainly for small sample problems. When the sample size $n > 1000$, the test statistics is impossible to handle (it requires inverting of an nxn matrix). Also, there is no tabulation for the distribution of the test statistics for large n.

The procedure adapted in this exercise is to utilize the Kolmogorov-Smirnov goodness-of-fit test to determine if one of the normal (or log-normal) distribution fit the samples. First, a brief description of the Kolmogorov-Smirnov test is given. The Kolmogorov-Smirnov test (see Ref. [3]) is a non-parametric statistical method for testing the hypothesis that a set of samples was drawn from a distribution, say F(x). The Kolmogorov-Smirnov statistic is defined by

$$D_n = \sqrt{n} \max | F(x) - F_n(x) | \qquad (I.1)$$

where $F_n(x) = k(x)/n$ and k(x) is the number of observations less than x.

The essence of this test is that, if the samples were drawn from F(x), the empirical distribution $F_n(x)$ should be close to F(x). So, if D_n is small within an acceptable bound, then the hypothesis is accepted. Using a 99% confidence level, this bound is 1.63. In other words, if D_n is greater than 1.63, then the hypothesis will be rejected.

In the present case, it requires finding the mean and standard deviation that minimize the Kolmogorov-Smirnov statistics. If this (minimum) statistic is less than 1.63, then the concentration is considered to have a log-normal distribution.

Using the Bruce NGS data (1981-1984), the null hypothesis is chosen to be a log-normal distribution with mean = 41.89$\mu Ci/m^3$ and standard deviation = 64.26$\mu Ci/m^3$. For the Pickering data, the hypothesized mean and standard deviation are 11.53$\mu Ci/m^3$ and 47.53$\mu Ci/m^3$, respectively.

Figures 1 and 2 show that the errors between the theoretical log-normal distribution of the null hypothesis and empirical distribution (of both the Pickering and the Bruce ALW discharge concentration) are very small. The Kolmogorov-Smirnov statistics indicated in the Figures are less than 1.63. Hence, the null hypothesis is accepted.

APPENDIX II

The Confidence Limit of an Estimated Probability

Let q = Probability that the concentration > 24000 $\mu Ci/m^3$. The value q depends on the mean, μ, and the standard deviation, σ, of the distribution of the log transforms. In other words, for a given set of (μ,σ), it determines a probability q, namely

$$q = P\{ \text{ concentration} > 24000\ \mu Ci/m^3 \mid \mu,\sigma \}$$

$$= 1 - \Phi[(10.086-\mu)/\sigma] \qquad \text{(II.1)}$$

where Φ is standard normal cumulative distribution. Note that $ln(24000) = 10.086$.

If μ and σ are known, then there will be no uncertainty. In this case, they are estimated and hence are random variables. Observe that the event $q < U$ occurs whenever (μ,σ) satisfies that inequality $1 - \Phi[(10.086-\mu/\sigma)] < U$. Hence,

$$P\{ q < U \} = P\{(\mu,\sigma) \mid 1 - \Phi[(10.086-\mu)/\sigma] < U\}$$

If $f(\sigma)$ and $g(\mu)$ are the probability density functions of σ and μ respectively, then

$$P\{ q < U \} = \int\int_{\Omega} f(\sigma)g(\mu)d\sigma d\mu \qquad \text{(II.2)}$$

where

$$\Omega = \{(\mu,\sigma) \mid 1 - \Phi([(10.086-\mu)/\sigma] < U\}$$

$$= \{(\mu,\sigma) \mid 10.086-\mu > \sigma\Phi^{-1}(1-U)\}$$

The one sided confidence interval of q, with confidence level 100α (eg $\alpha = .99$) is defined by the value U such that

$$P\{ q < U \} = \alpha$$

or

$$\alpha = \int_{-\infty}^{\infty} [\int_{0}^{h(\mu,U)} f(\sigma)d\sigma\ ]g(\mu)d\mu \qquad \text{(II.3)}$$

where $h(\mu,U) = (10.086-\mu)/\Phi^{-1}(1-U)$.

The distribution $g(\mu)$ can be assumed to be a normal distribution with mean μ_n and standard deviation $\sigma_n/\sqrt{n}$ where μ_n is the sample mean and σ_n is the sample standard deviation. Since n is large, the value $\sigma/\sqrt{n}$ is very small. Hence, the mean μ can be considered as a constant equal to μ_n.

The distribution $f(\sigma)$ can be derived as follows:

Let S denote the random variable of the standard deviation, ie

$$P\{ S < \sigma \} = \int_0^{\sigma} f(s)ds$$

It is well known that $(n-1)S^2/\gamma^2$ is a chi-square distribution with n-1 degree of freedom, where γ^2 is the variance of the population. Thus,

$$P\{ S < \sigma \}$$

$$= P\{ (n-1)S^2/\gamma^2 < (n-1)\sigma^2/\gamma^2 \}$$

$$= \int_0^{(n-1)\sigma^2/\gamma^2} C_n t^{(n-3)/2}\exp\{-t/2\}dt$$

where $C_n = 2^{-(n-1)/2} \int_0^{\infty} s^{(n-3)/2}\exp\{-s/2\}ds$

Now, one can re-write (II.3) as

$$\alpha = \int_0^{G(\mu,n,\gamma)} f_{n-1}(s)ds$$

where f_{n-1} is the Chi-square distribution with n-1 degree of freedom and

$$G(\mu,n,\gamma) = (n-1)[(10.086 - \mu)/\gamma\Phi^{-1}(1-U)]^2 \qquad (II.4)$$

Hence, one can write

$$G(\mu,n,\gamma) = F^{-1}_{n-1}(\alpha)$$

where $F^{-1}_{n-1}(\alpha)$ is the inverse of the cumulative Chi-square distribution with n-1 degrees of freedom.

Now, consider the following approximation

$$F_n^{-1}(x) \cong n(1 - 0.2222/n + \Phi^{-1}(x)(0.2222/n)^{1/2})^3 \quad (II.5)$$

where F_n^{-1} and Φ^{-1} are the inverses of the Chi-square (with n degrees of freedom) and the normal cumulative distributions, respectively. Substituting (II.4) and (II.5) into the above equation and using $\Phi^{-1}(0.99) \cong 2$, one can obtain

$$U = 1 - \Phi[(10.086 - \mu)/\theta\gamma] \quad (II.6)$$

where

$$\theta = (1 - 0.2222/(n-1) + 2(0.2222/(n-1))^{1/2})^{3/2}$$

Now, μ can be substituted by the sample mean. But if the value of γ is substituted by the sample deviation, the confidence interval will be very tight. Therefore, in order to get a more conservative estimate, it will be reasonable to increase the standard deviation by 20%. Hence, γ will be substituted by $1.2\sigma_n$, where σ_n is the sample standard deviation.

It must be noted here that the sample mean and standard deviation are for the log transform of the concentrations. Their relationship is defined by the following equations:

$$\text{Sample Mean} = \exp\{\mu + \sigma^2/2\} \quad (II.7)$$

$$\text{Sample S.D.} = \exp\{\mu + \sigma^2/2\}(\exp\{\sigma^2\}-1)^{1/2} \quad (II.8)$$

where μ and σ are the mean and standard deviation of the log transforms, respectively.

Using the Bruce NGS data (1981 - 1984) and the Pickering NGS data (1982 - 1984) the following results are obtained:

(1) The sample mean and sample standard deviation of the Bruce ALW data are 45.69μCi/m^3 and 78.62μCi/m^3 respectively. Using (II.1),(II.6),(II.7) and (II.8), the probability that the concentration of a discharge is greater than 24000μCi/m^3 is

$$q = 1 - \Phi(5.92)$$

$$\leq 10^{-8}$$

$$U = 1 - \Phi(4.8)$$

$$= 7.2 \times 10^{-7}$$

(2) For the Pickering ALW data, the sample mean and standard deviation are $13.14 \mu Ci/m^3$ and $59.31 \mu Ci/m^3$ respectively. From this,

$$q \leq 10^{-8} \quad *$$

and

$$U = 1 - \Phi(4.19)$$

$$= 1.4 \times 10^{-5}$$

* *These values are below the approximation error.*

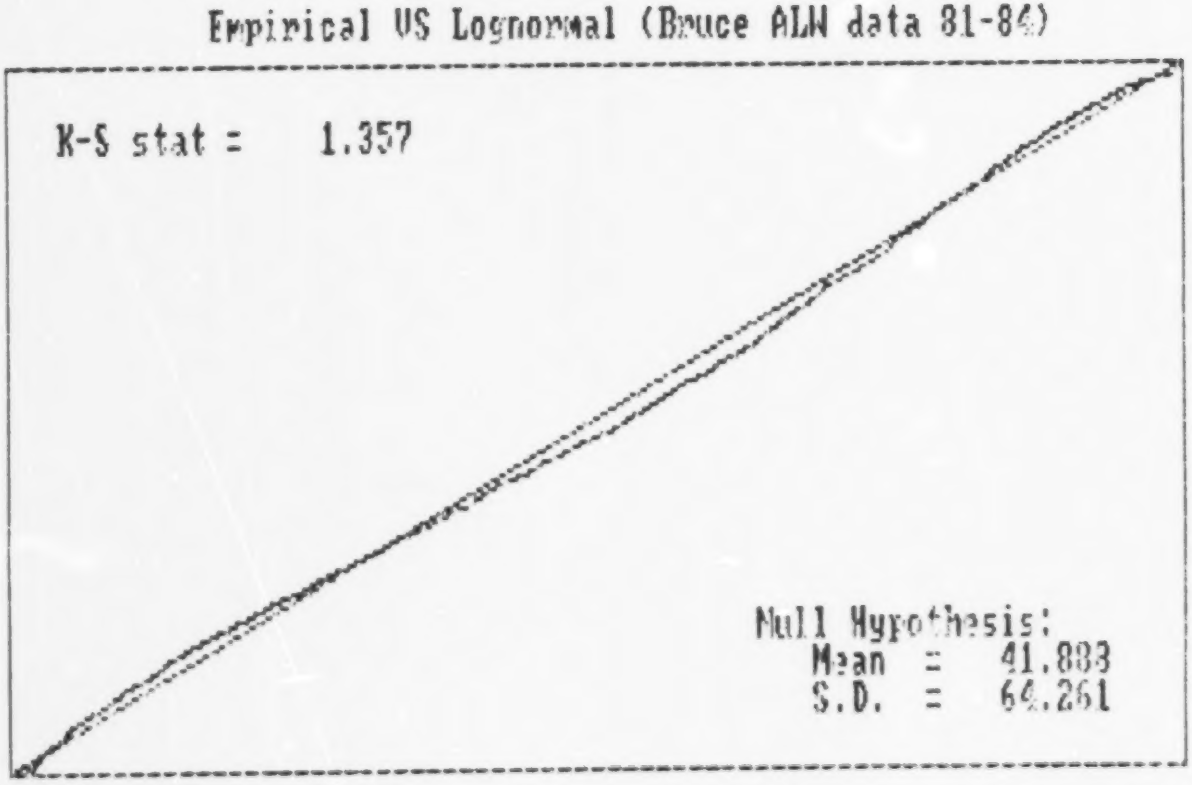

Figure 1

Sample Distribution of the Bruce NGSA ALW
Discharge Concentration vs a Log-normal Distribution

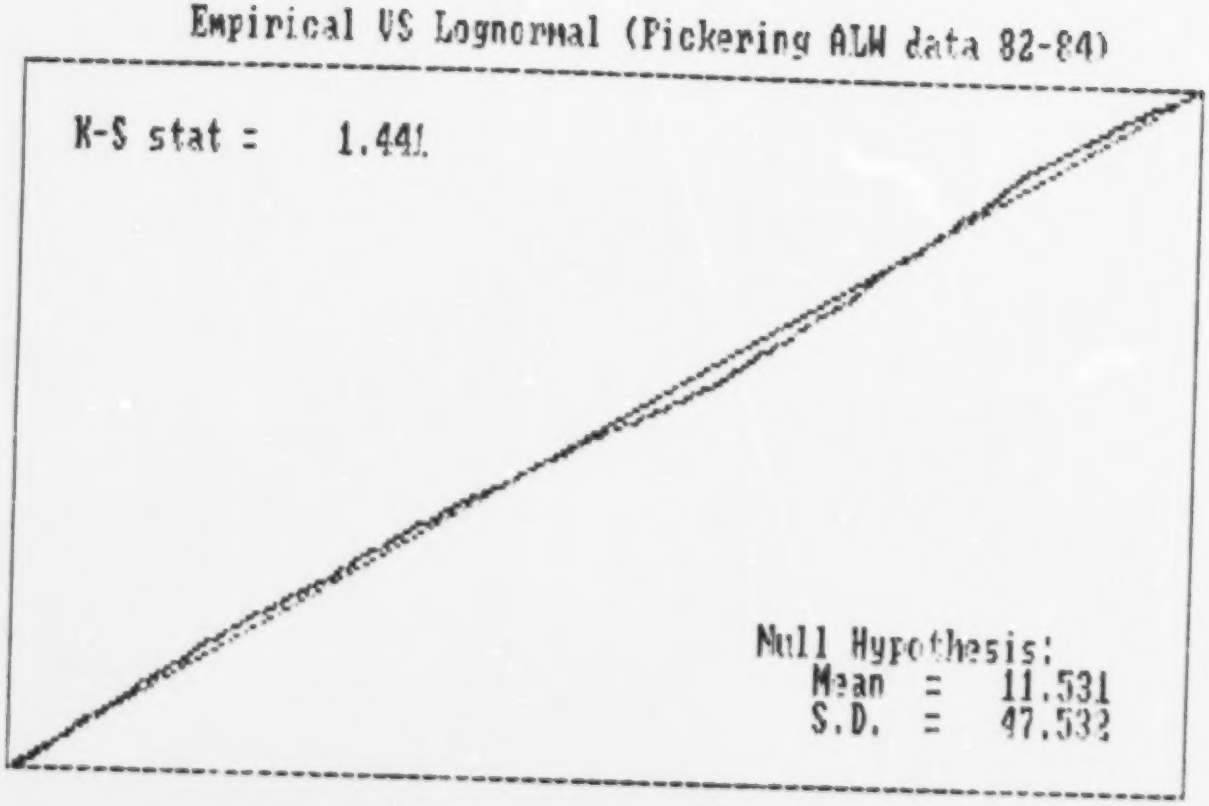

Figure 2

Sample Distribution of the Pickering NGSA ALW
Discharge Concentration vs a Log-normal Distribution